Quantile regression in Stata

Performance, precision, and power

Morten Wang Fagerland

Oslo Centre for Biostatistics and Epidemiology
Oslo University Hospital
Norway

Sample size: 50 in each group (quantile = 75%)

VCE*	DET†	Method	$\gamma = 0$	$\gamma = 2.9$	$\gamma = 4.0$	$\gamma = 4.8$	$\gamma = 8.9$
iid	fitted	hsheather	4.3	3.2	2.7	2.9	0.18
		bofinger	3.8	2.4	1.8	2.1	0.08
		chamberlain	7.1	6.5	6.3	6.3	1.7
	residual	hsheather	3.8	3.5	4.1	3.4	1.0
		bofinger	3.3	2.6	2.9	2.5	0.29
		chamberlain	5.8	6.2	8.2	6.5	10.8
	kernel	epanechnikov	1.2	7.3	27.7	10.3	70.5
		epan2	3.9	8.3	24.8	9.7	72.1
		biweight	4.6	8.4	21.3	9.6	70.3
		cosine	7.1	10.1	16.7	11.4	59.7
		gaussian	1.4	6.8	25.6	9.5	69.5
		parzen	5.6	8.8	18.3	10.0	66.2
		rectangle	3.0	8.4	33.5	10.5	71.2
		triangle	4.5	8.7	23.5	10.0	70.2
robust	fitted	hsheather	4.1	2.8	2.1	2.6	0.08
		bofinger	3.6	2.2	1.4	1.8	0.04
		chamberlain	6.6	5.7	5.1	5.4	0.50
	kernel	epanechnikov	1.2	6.2	25.0	8.7	68.5
		epan2	3.5	5.3	7.9	5.6	0.78
		biweight	4.1	5.3	8.0	5.6	1.6
		cosine	5.7	7.2	9.9	7.8	9.9
		gaussian	1.3	5.5	19.9	7.3	61.4
		parzen	4.7	6.0	8.8	6.4	3.9
		rectangle	2.6	5.4	7.6	5.7	0.40
		triangle	4.0	5.7	8.7	6.0	1.7
-	-	bootstrap 20‡	5.6	5.1	4.7	4.7	0.61
-	-	bootstrap 100‡	4.2	3.8	3.2	3.4	0.09

*Variance-covariance estimator
†Density estimation technique
‡Number of bootstrap replications

QUANTILE REGRESSION

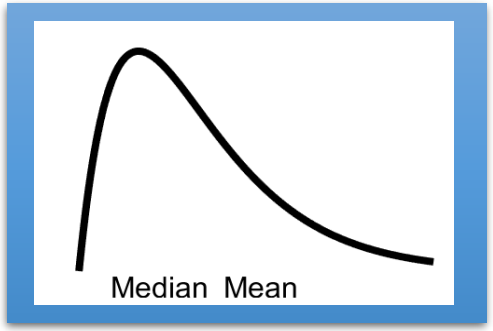
- Quantile regression **estimates quantiles of the outcome variable**, conditional on the values of the independent variables, with median regression as the default form
- Stata command: `qreg`
- Method of minimum **absolute deviations**

MOTIVATION

- Median regression as an **alternative to linear regression**
  - estimate medians instead of means as a measure of central tendency

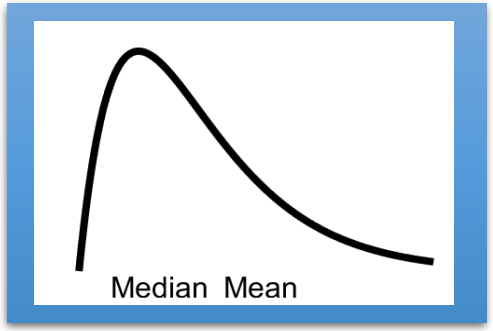
Potentially useful when mean ≠ typical observation

Potentially useful when mean ≠ typical observation

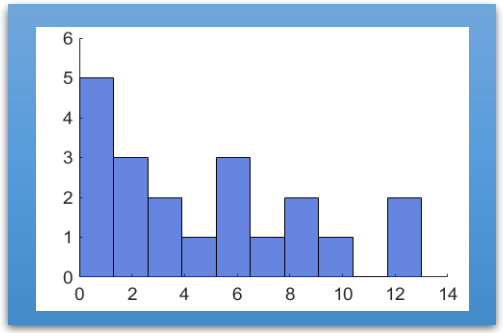


When data are markedly skewed

Potentially useful when mean ≠ typical observation

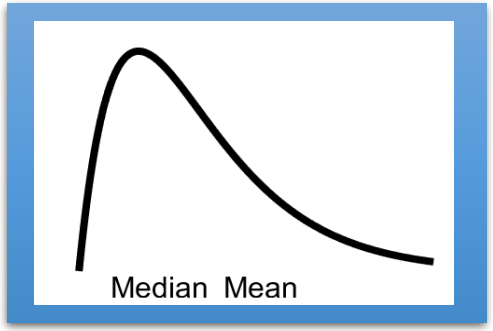


When data are markedly skewed

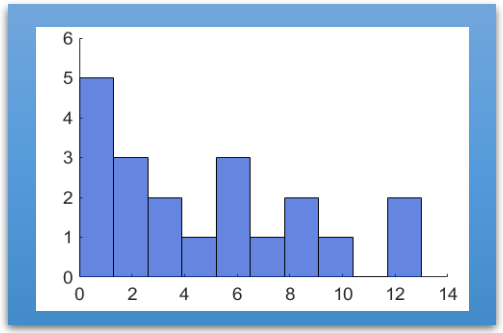


If sample is small → hard to assess distribution

Potentially useful when mean ≠ typical observation



When data are markedly skewed



If sample is small → hard to assess distribution

RCT

Statistical Analysis Plan

If methods need to be pre-specified & distribution is unknown

APPLICABLE TO QUANTILE REGRESSION IN GENERAL

- To estimate a **particular quantile of interest**, such as the 10th quantile of birth weight to find predictors of low birth weight
- To study how the **effect of independent variables vary over different quantiles** of the dependent variable

PROBLEM

- How to specify the variance-covariance estimator?
- So many to choose from: 26!
- And they can give quite different results

PROBLEM

- How to specify the variance-covariance estimator?
- So many to choose from: 26!
- And they can give quite different results

WHY?

WHICH?

PROBLEM

- How to specify the variance-covariance estimator?
- So many to choose from: 26!
- And they can give quite different results

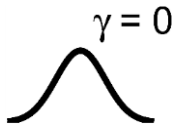
~~WHY?~~

WHICH?

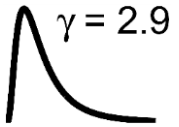
AIM

- To explore the **performance** of the methods and to arrive at some overall **recommendations** for which methods to use.

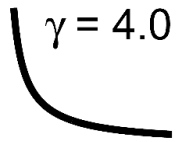
SIMULATIONS



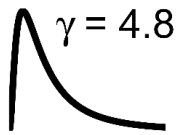
Standard normal distribution



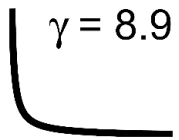
Lognormal distribution with μ = 0, σ = 0.7



Gamma distribution with shape = 0.25 and scale = 4.0



Lognormal distribution with μ = 0, σ = 0.9



Gamma distribution with shape = 0.05 and scale = 50

SIMULATIONS (cont)

- **Replications**: 10 000
- **Precision** (expected): for π=5%: 95% CI 4.57% to 5.43%
- **Sample sizes**: 25, 50, 100, 500 in each of two groups
- **Quantiles**: 50%, 75%, 90%
- **Covariates**: no, weakly, moderately, or strongly correlated to outcome

PERFORMANCE MEASURES

- **Rejection rates** of significance test
- **Power** of test
- **Coverage probability** of confidence interval
- **Width** of confidence interval

- *(**Location** of confidence interval)*

PERFORMANCE MEASURES

Rejection rate equivalent to coverage probability

- **Rejection rates** of significance test
- **Power** of test
- **Coverage probability** of confidence interval
- **Width** of confidence interval

- *(**Location** of confidence interval)*

PERFORMANCE MEASURES

- **Rejection rates** of significance test
- **Power** of test
- **Coverage probability** of confidence interval
- **Width** of confidence interval

- *(**Location** of confidence interval)*

Rejection rate
equivalent to
coverage probability

Power
equivalent to
width

PERFORMANCE MEASURES

- **Rejection rates** of significance test
- **Power** of test
- **Coverage probability** of confidence interval
- **Width** of confidence interval

- *(**Location** of confidence interval)*

Rejection rate
equivalent to
coverage probability

Power
equivalent to
width

Rejection rate
more important than
Power

PERFORMANCE MEASURES

- **Rejection rates** of significance test
- **Power** of test
- **Coverage probability** of confidence interval
- **Width** of confidence interval

Start looking at rejection rates

Rejection rate equivalent to coverage probability

Power equivalent to width

Rejection rate more important than Power

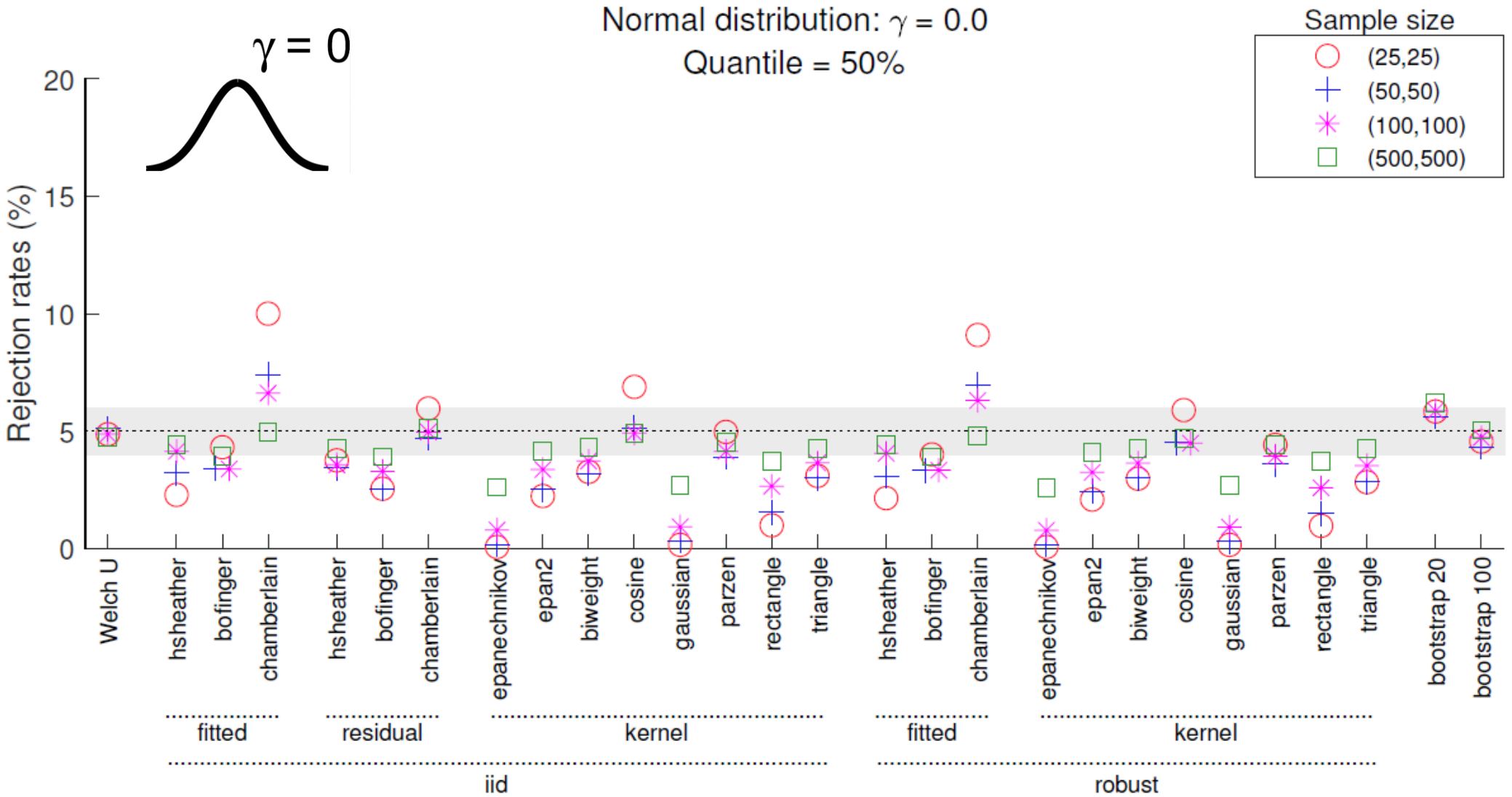
VCE	DET	Bandwidth	Kernel function	Option to qreg
iid	fitted	Hall-Sheather	-	`vce(iid, fitted hsheather)`*
iid	fitted	Bofinger	-	`vce(iid, fitted bofinger)`
iid	fitted	Chamberlain	-	`vce(iid, fitted chamberlain)`
iid	residual	Hall-Sheather	-	`vce(iid, residual hsheather)`
iid	residual	Bofinger	-	`vce(iid, residual bofinger)`
iid	residual	Chamberlain	-	`vce(iid, residual chamberlain)`
iid	kernel	-	epanechnikov	`vce(iid, kernel(epanechnikov))`
iid	kernel	-	epan2	`vce(iid, kernel(epan2))`
iid	kernel	-	biweight	`vce(iid, kernel(biweight))`
iid	kernel	-	cosine	`vce(iid, kernel(cosine))`
iid	kernel	-	gaussian	`vce(iid, kernel(gaussian))`
iid	kernel	-	parzen	`vce(iid, kernel(parzen))`
iid	kernel	-	rectangle	`vce(iid, kernel(rectangle))`
iid	kernel	-	triangle	`vce(iid, kernel(triangle))`
robust	fitted	Hall-Sheather	-	`vce(robust, fitted hsheather)`
robust	fitted	Bofinger	-	`vce(robust, fitted bofinger)`
robust	fitted	Chamberlain	-	`vce(robust, fitted chamberlain)`
robust	kernel	-	epanechnikov	`vce(robust, kernel(epanechnikov))`
robust	kernel	-	epan2	`vce(robust, kernel(epan2))`
robust	kernel	-	biweight	`vce(robust, kernel(biweight))`
robust	kernel	-	cosine	`vce(robust, kernel(cosine))`
robust	kernel	-	gaussian	`vce(robust, kernel(gaussian))`
robust	kernel	-	parzen	`vce(robust, kernel(parzen))`
robust	kernel	-	rectangle	`vce(robust, kernel(rectangle))`
robust	kernel	-	triangle	`vce(robust, kernel(triangle))`
Bootstrap with # replications				`reps(#)`†

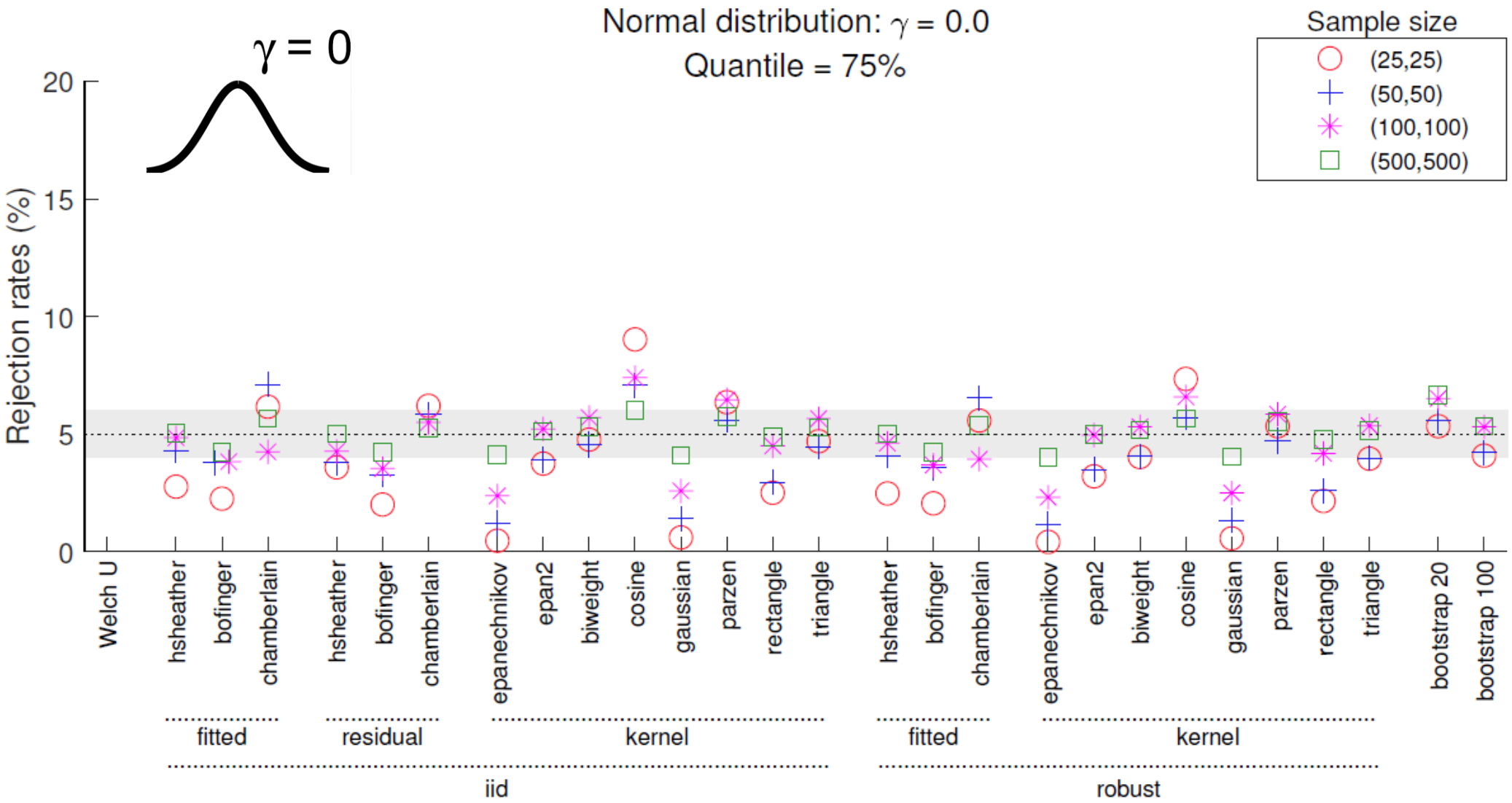
*Default method

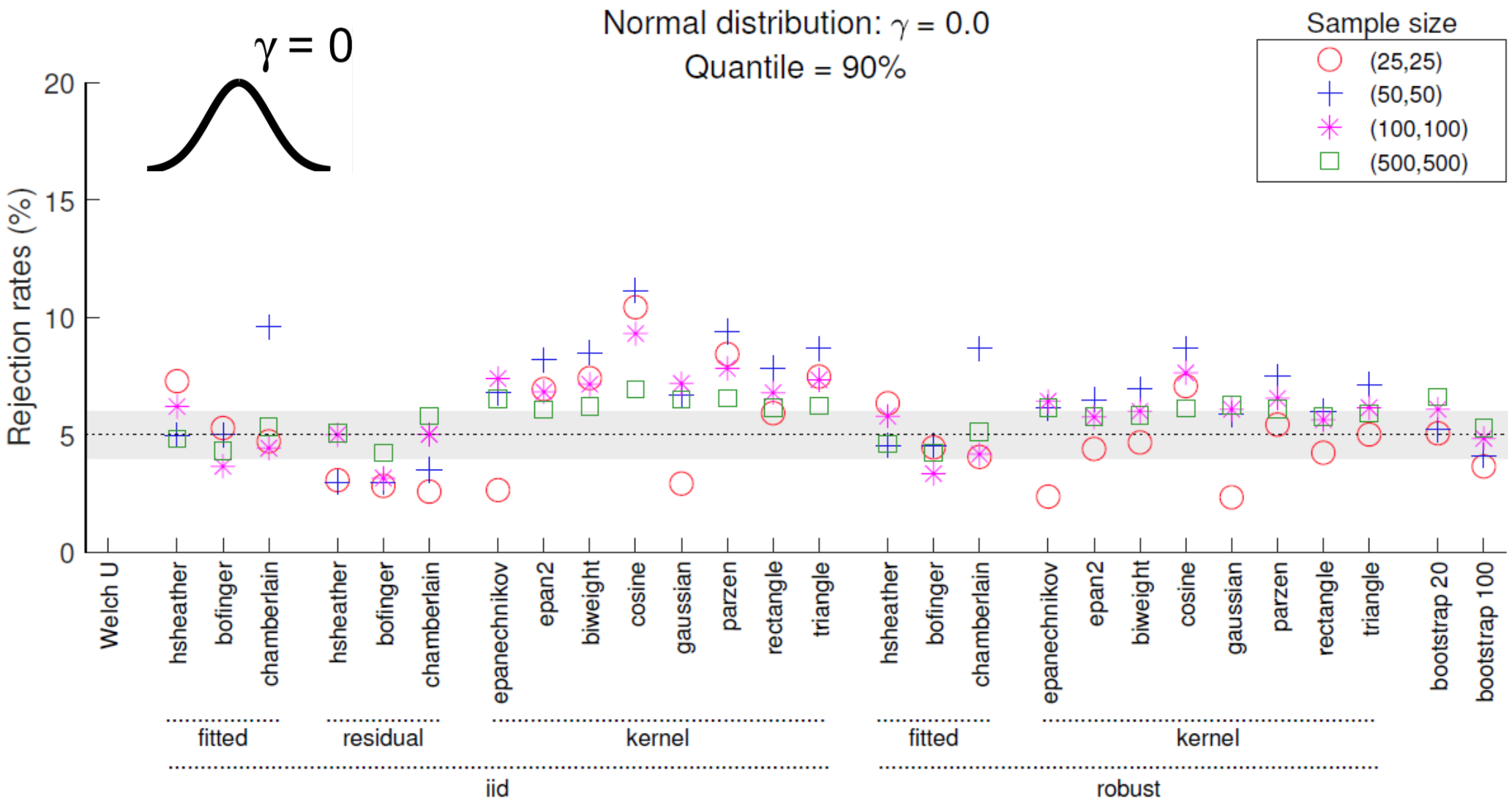
†Use the `sqreg` command instead of `qreg`

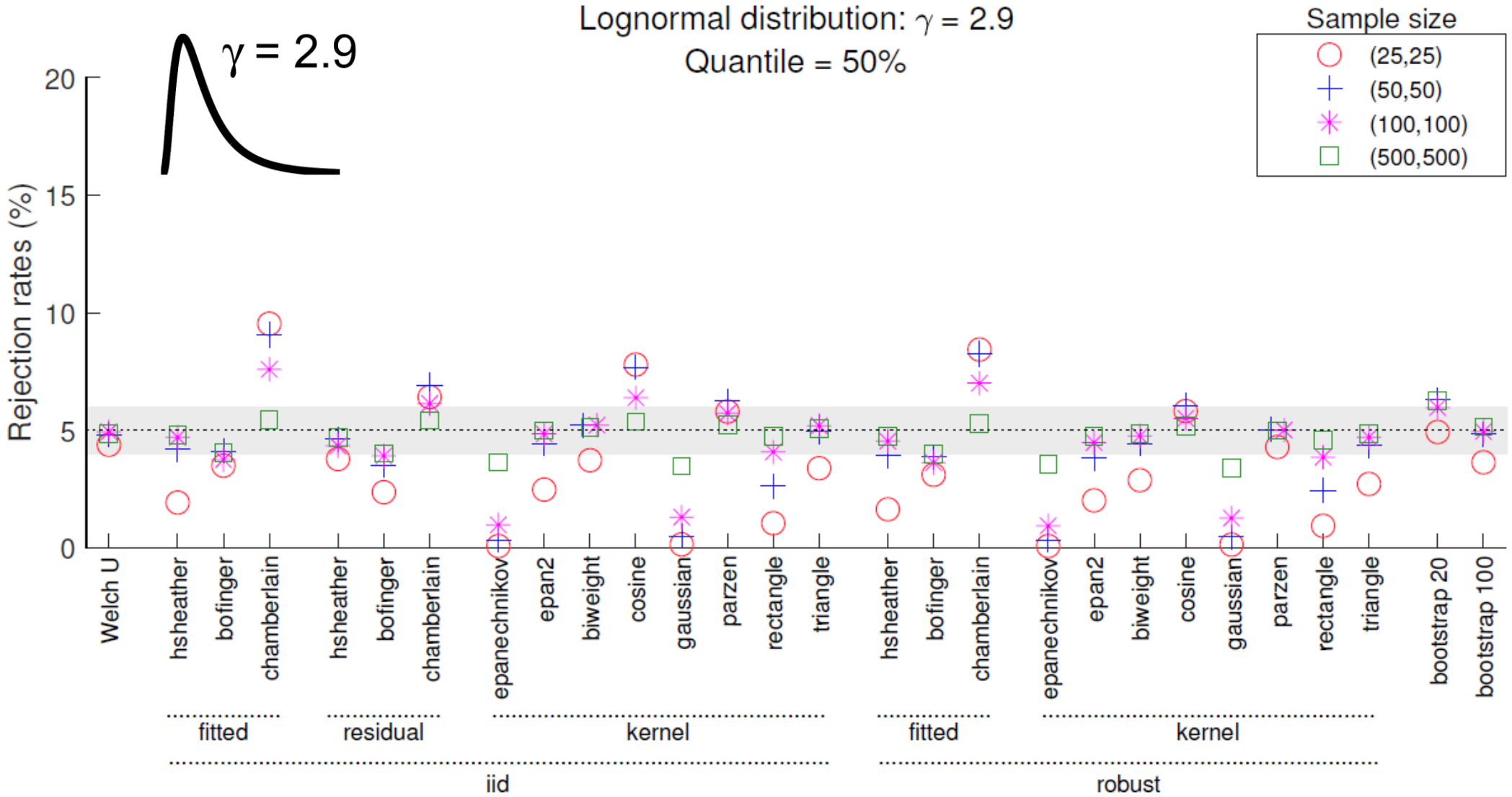
The variance-covariance estimators

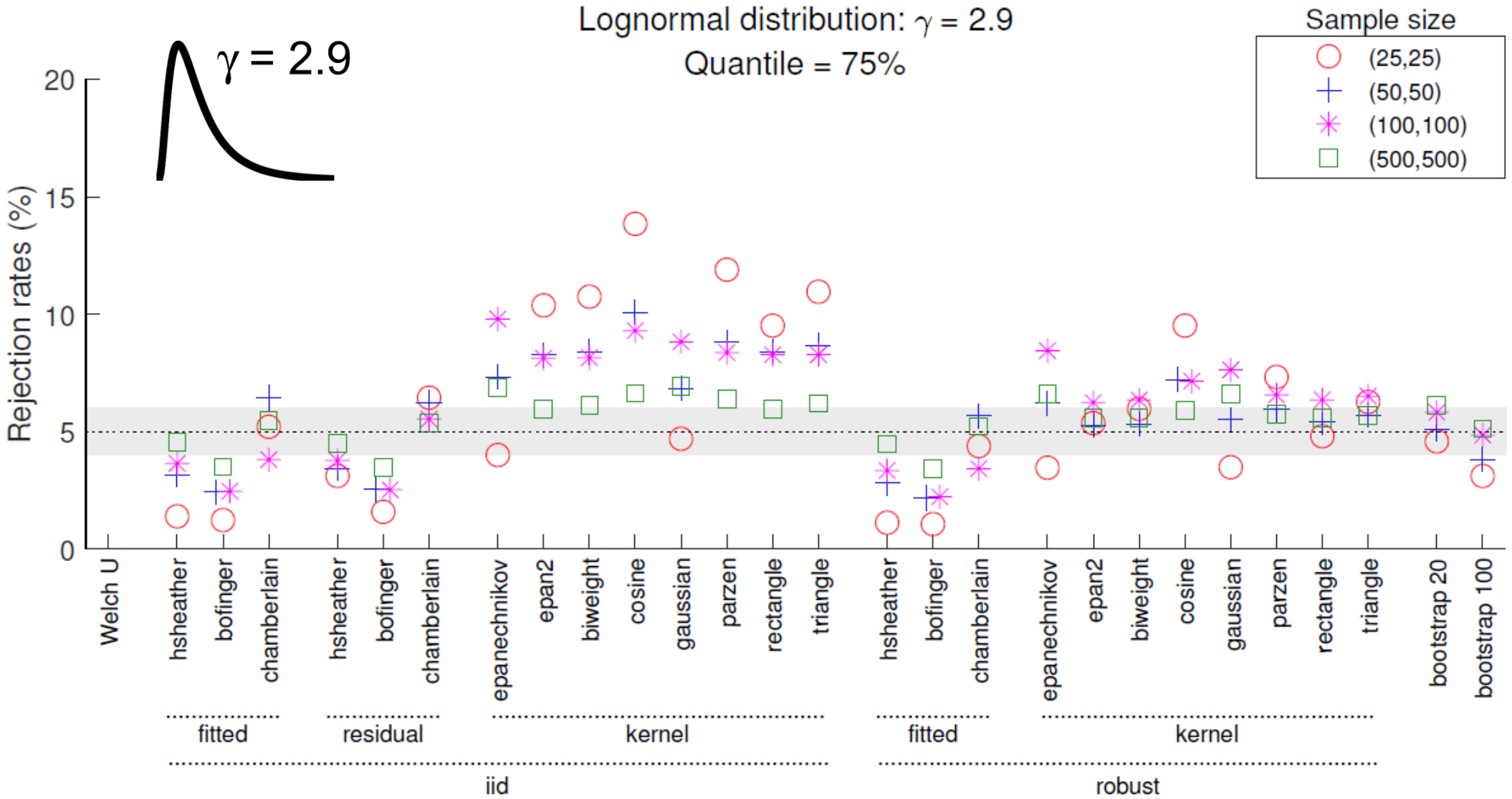
VCE = Variance-covariance estimator
DET = Density estimation technique
iid = independent & identically distributed

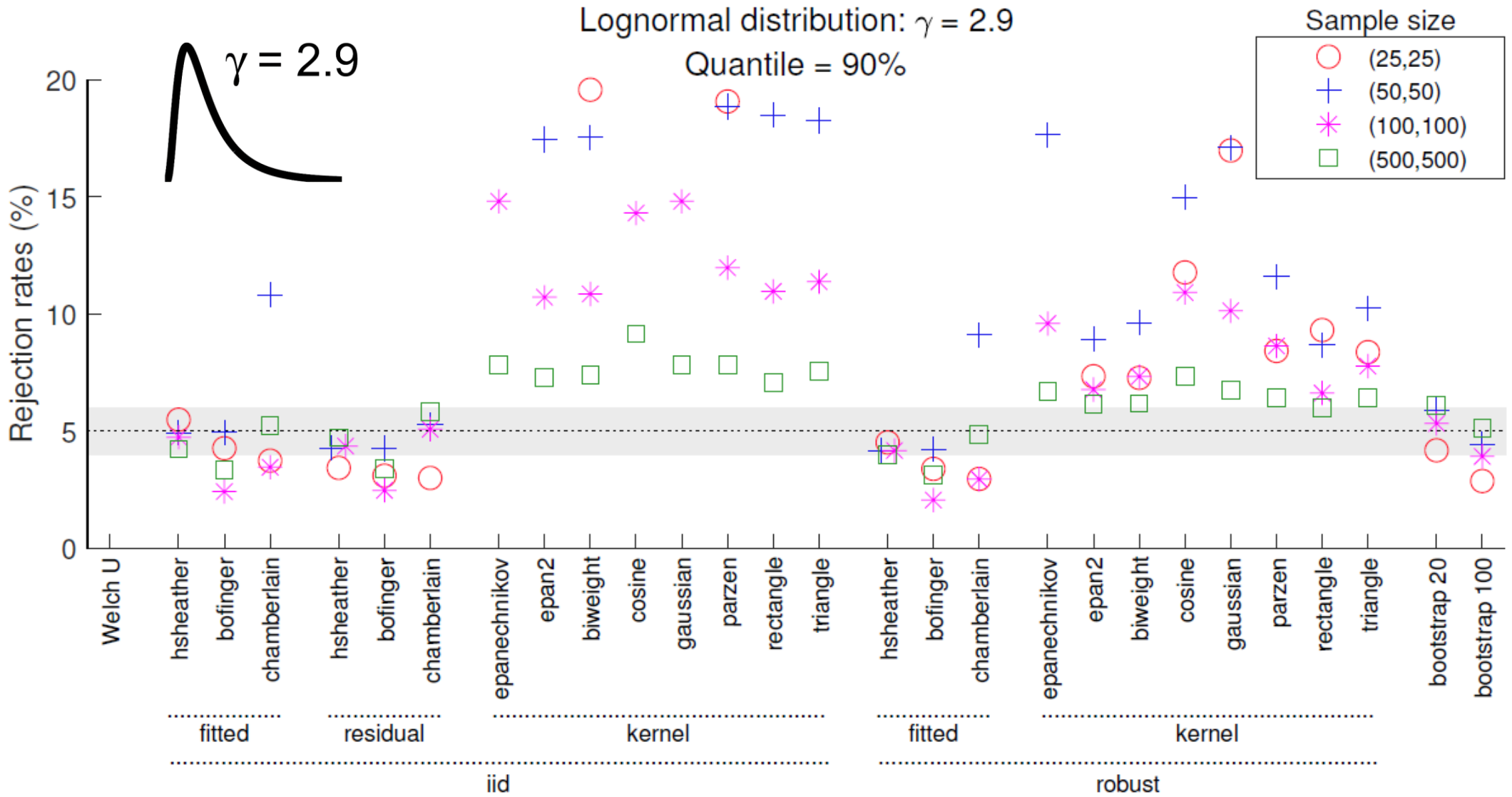

Normal distribution: γ = 0.0
Quantile = 50%
γ = 0
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

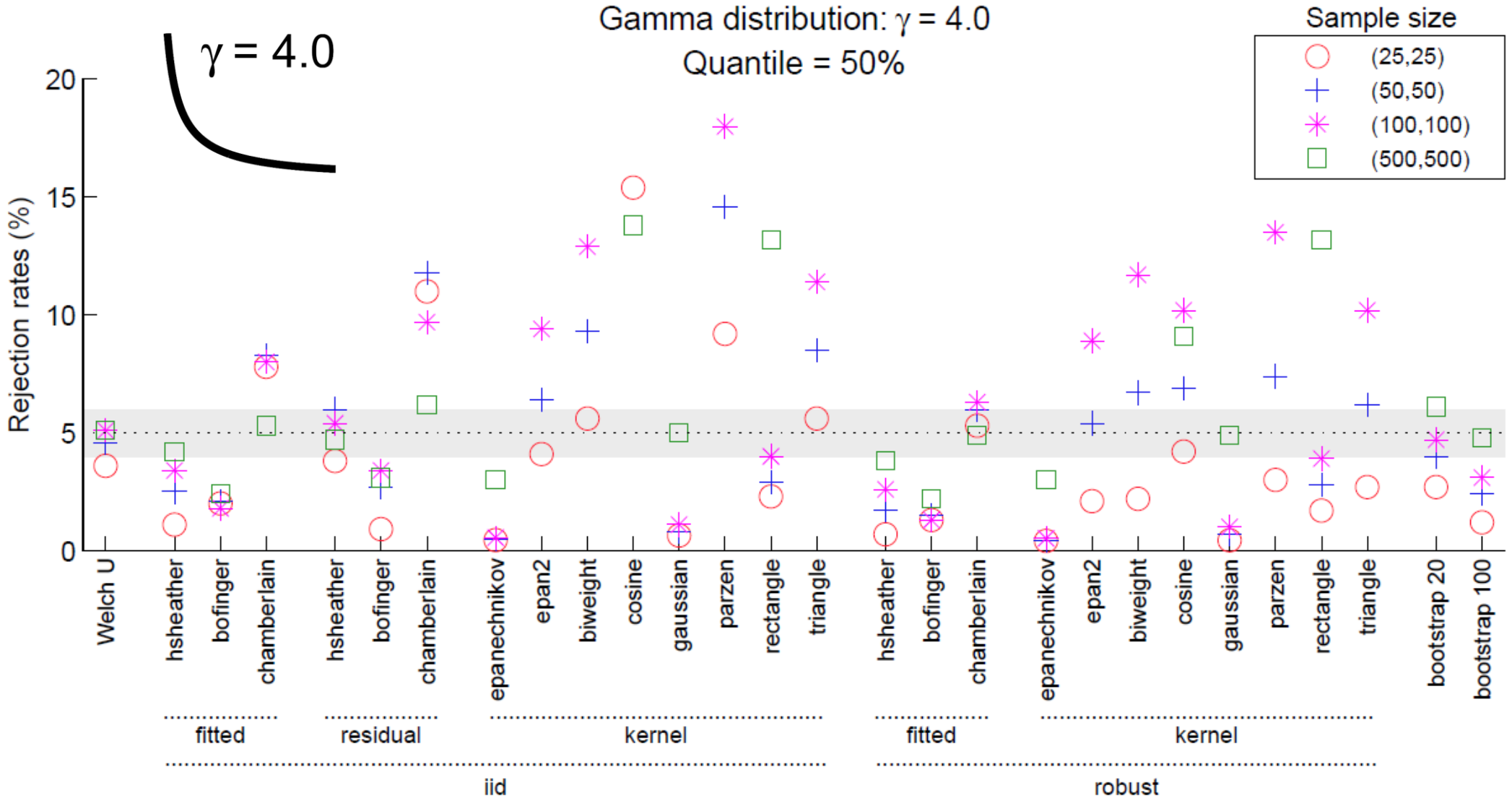

Normal distribution: γ = 0.0
Quantile = 75%
γ = 0
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

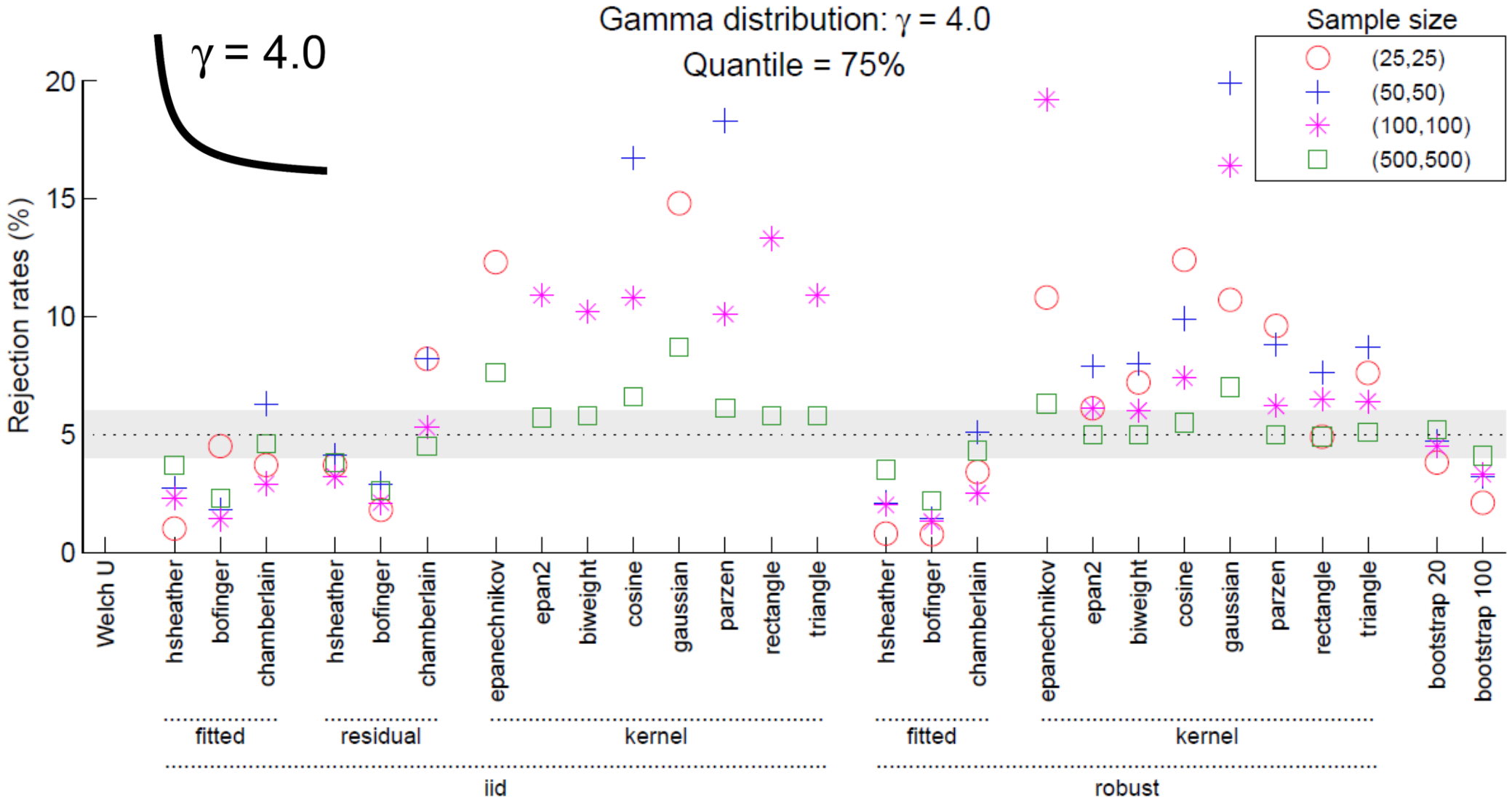

γ = 0
Normal distribution: γ = 0.0
Quantile = 90%
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

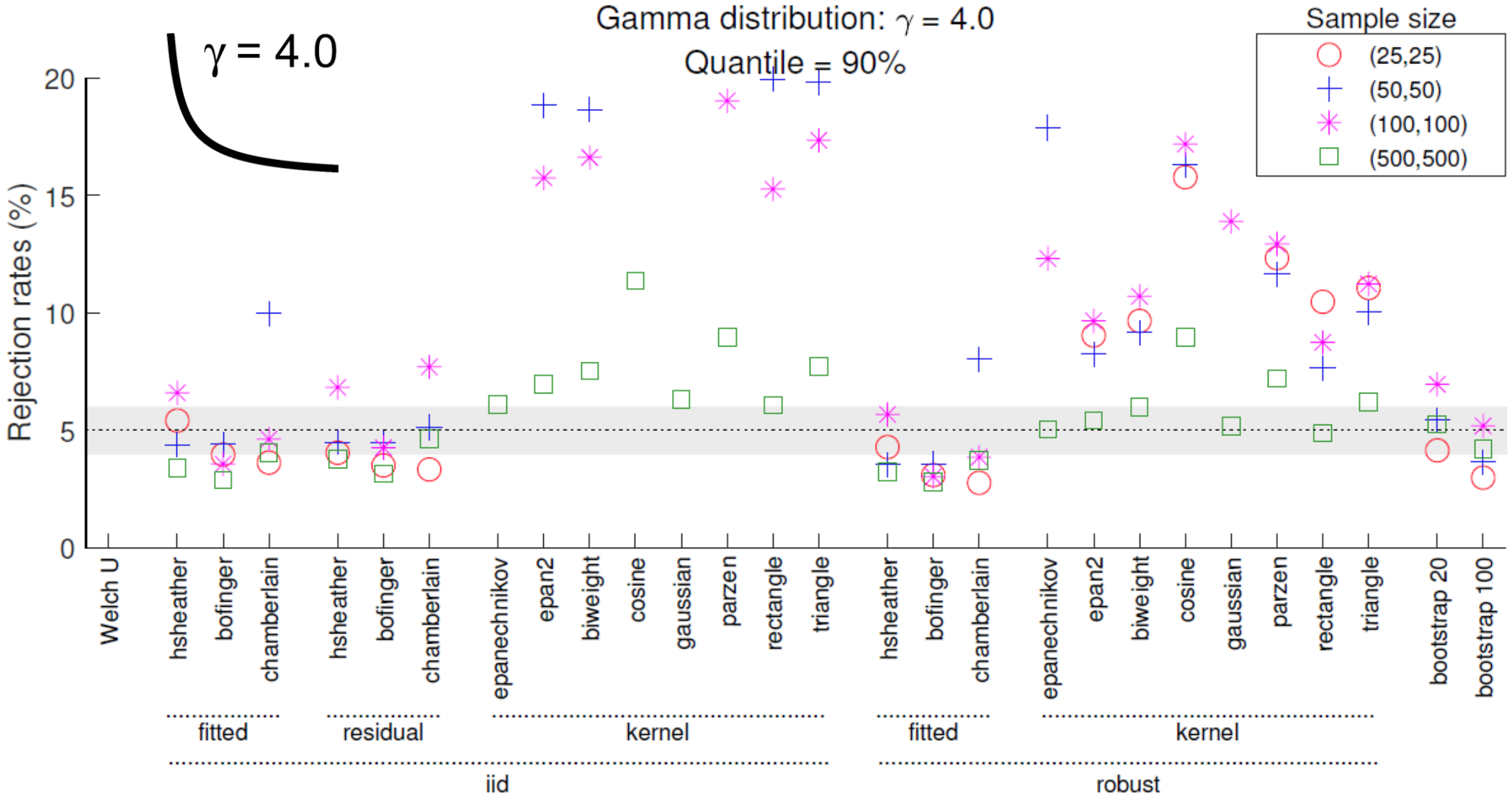

Lognormal distribution: γ = 2.9
Quantile = 50%
γ = 2.9
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

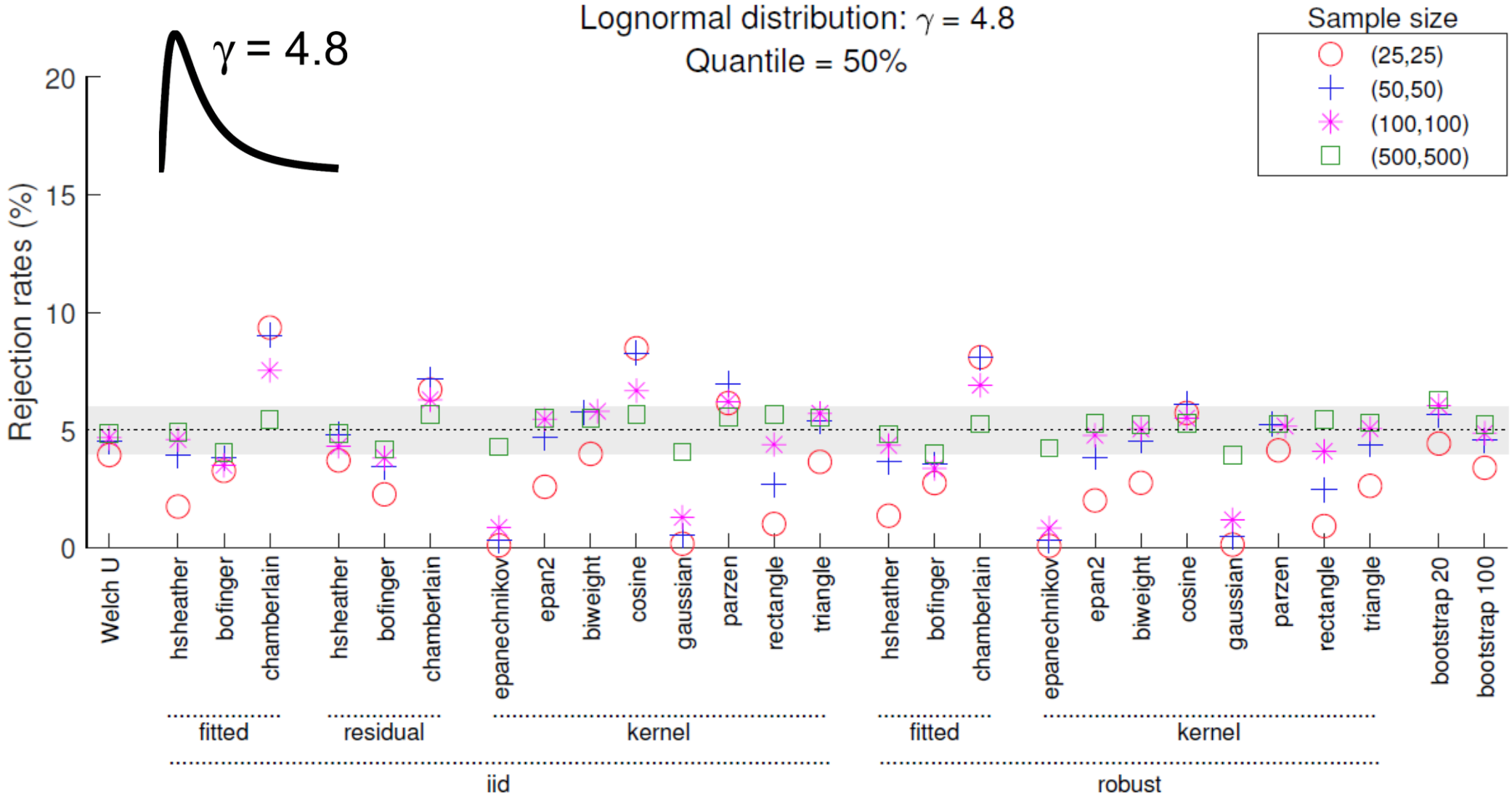

Lognormal distribution: γ = 2.9
Quantile = 75%
γ = 2.9
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
0
5
10
15
20
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

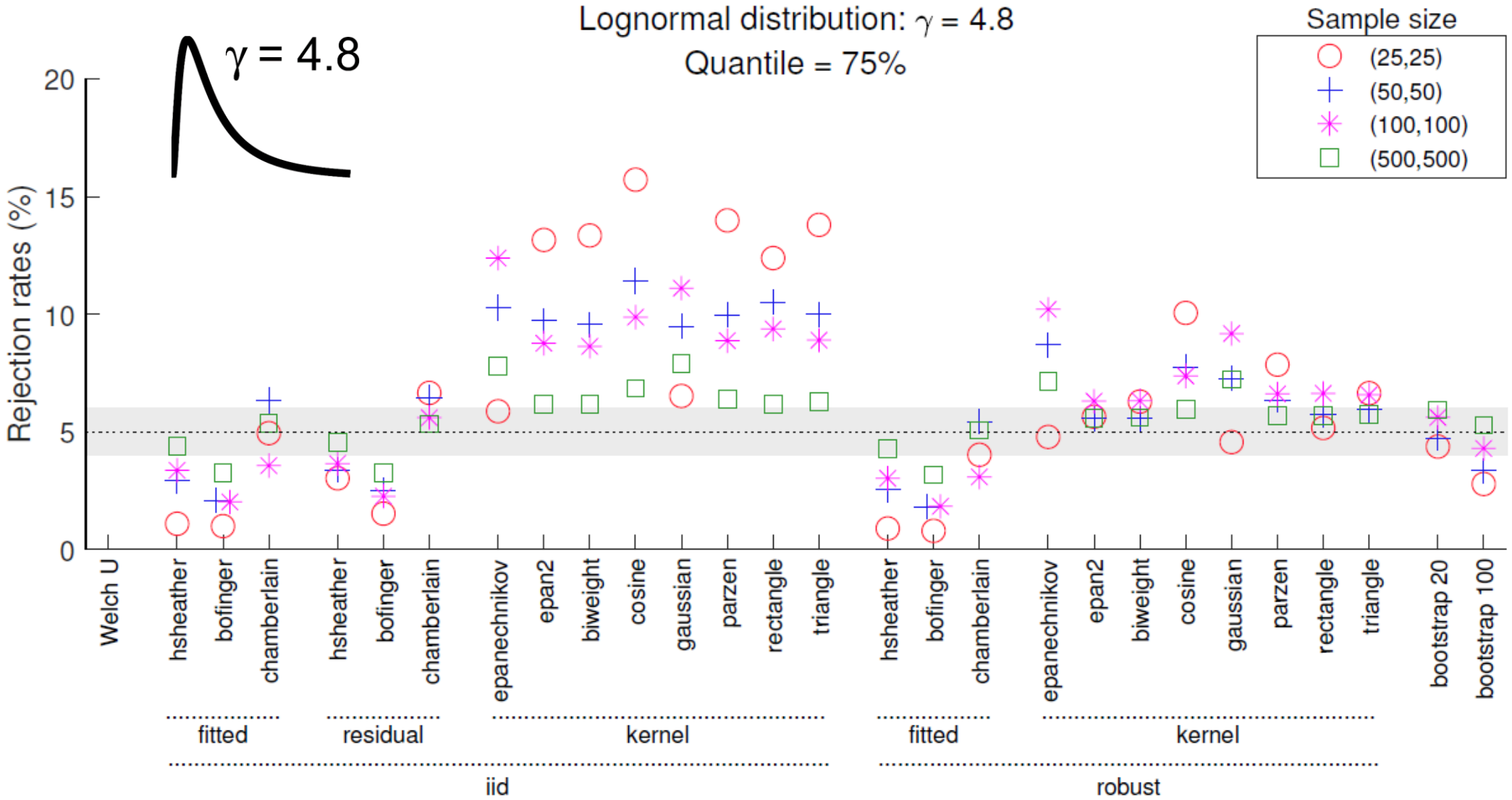

Lognormal distribution: γ = 2.9
Quantile = 90%
γ = 2.9
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
0
5
10
15
20
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

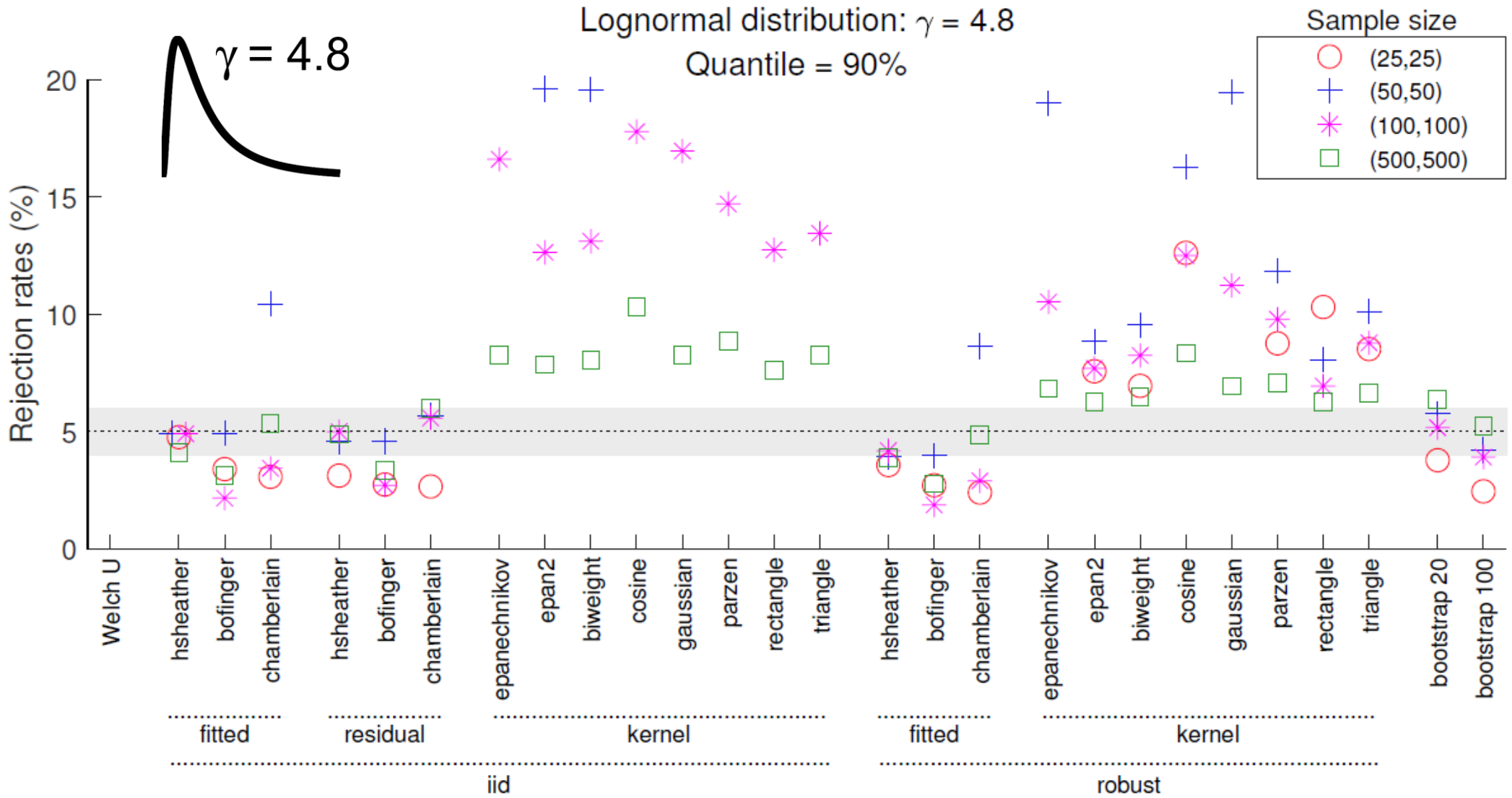

Gamma distribution: γ = 4.0
Quantile = 50%
γ = 4.0
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

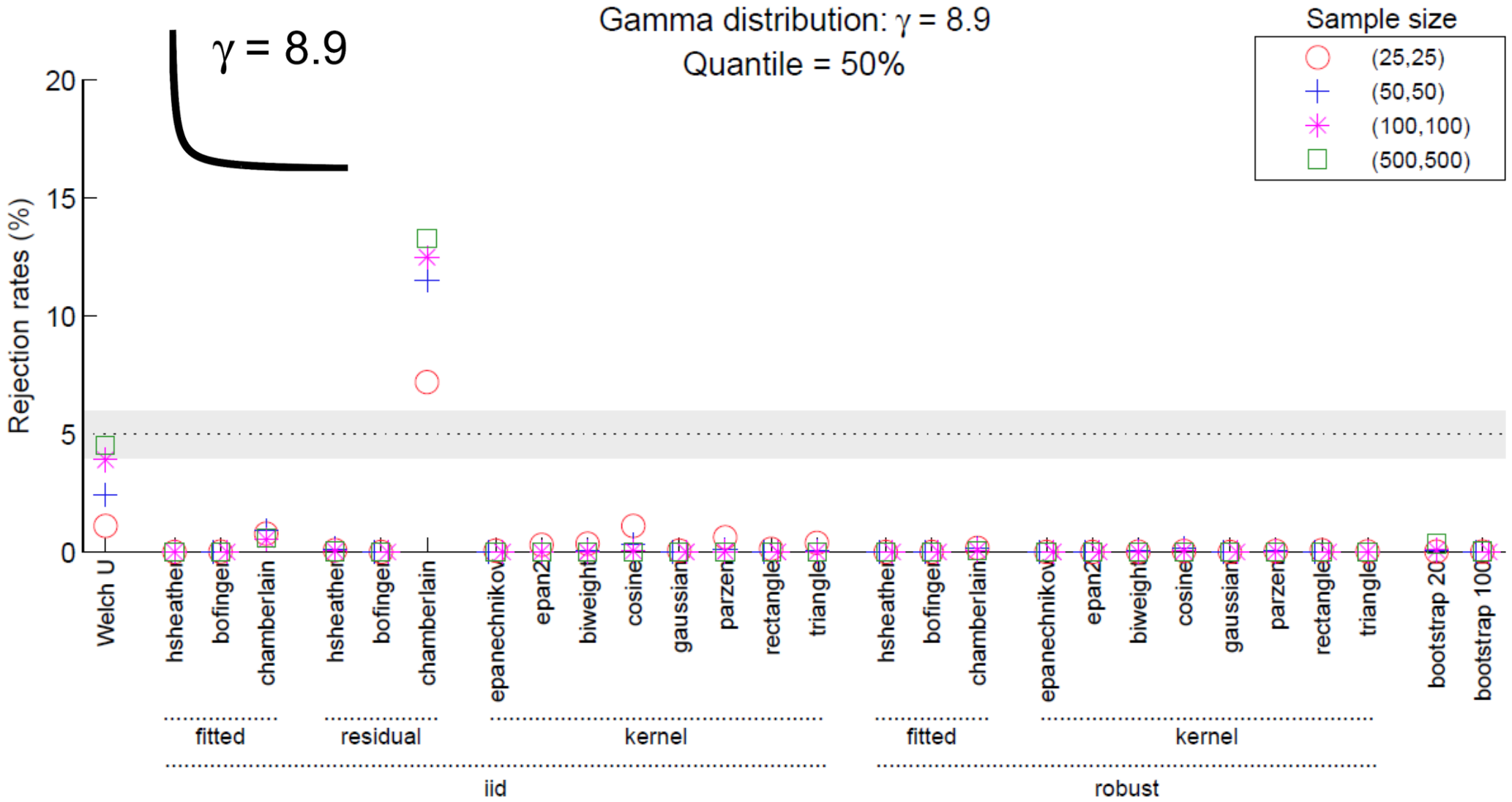

Gamma distribution: γ = 4.0
Quantile = 75%
γ = 4.0
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

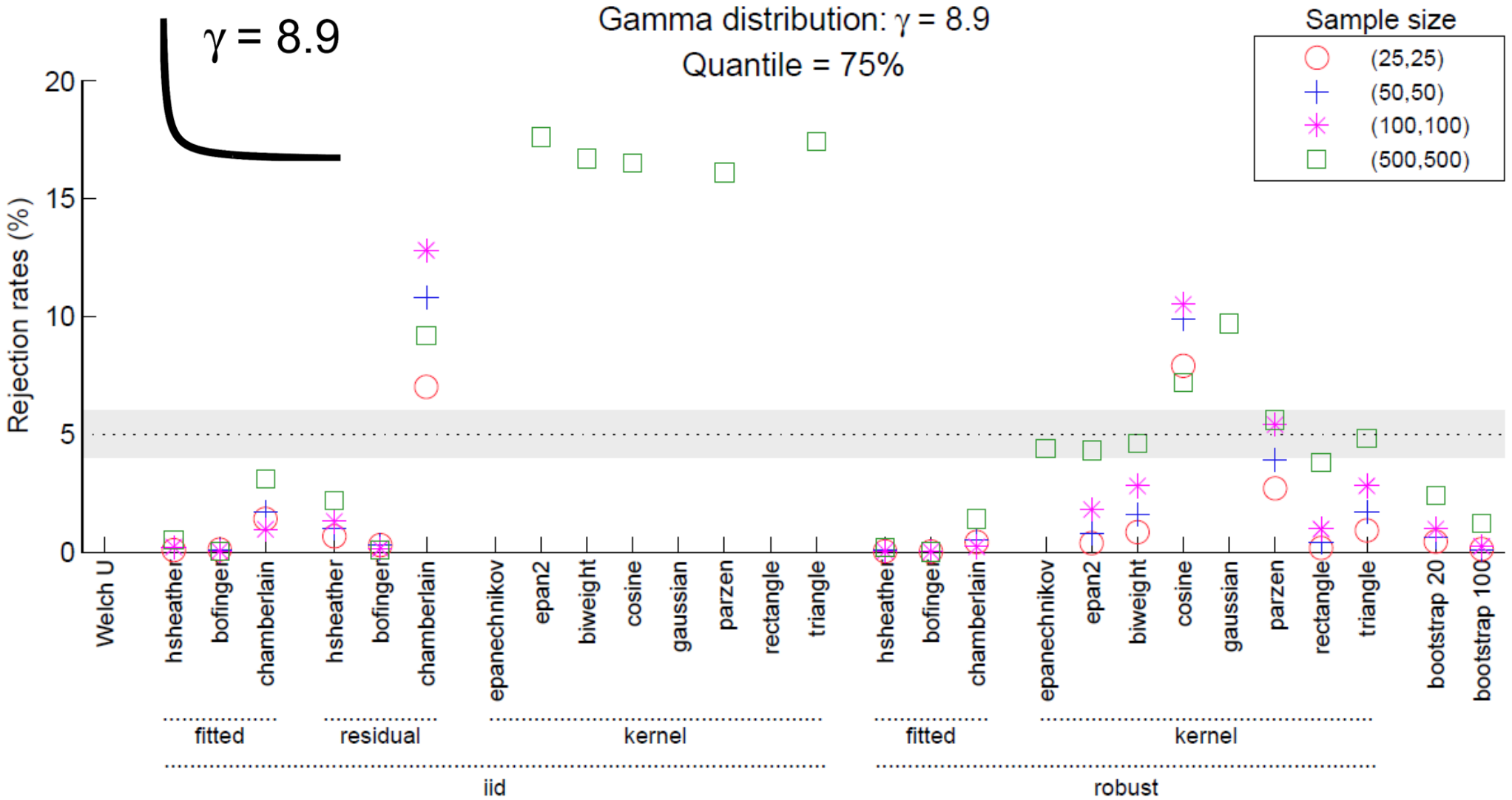

Gamma distribution: γ = 4.0
Quantile = 90%
γ = 4.0
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
0
5
10
15
20
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

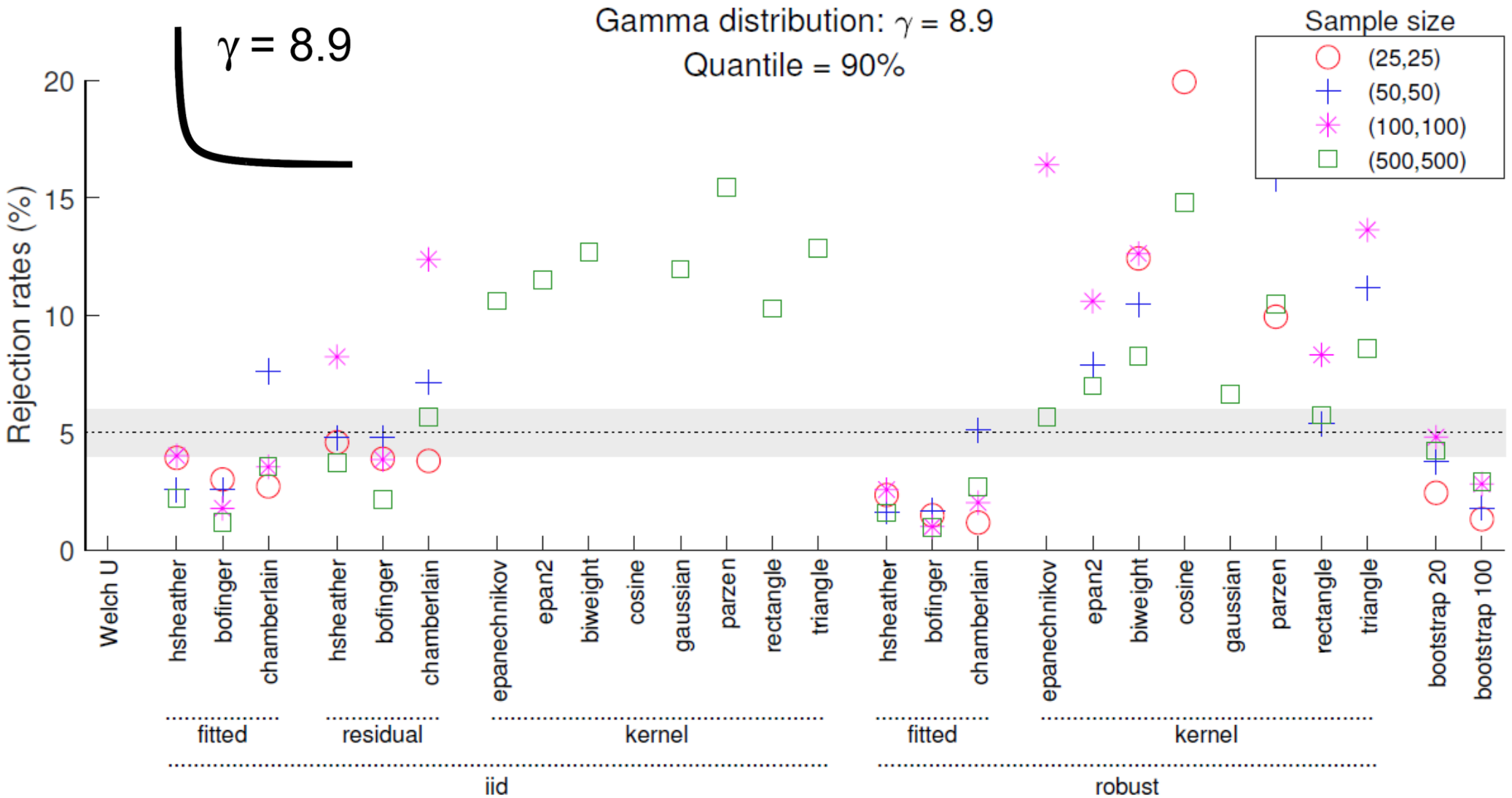

Lognormal distribution: γ = 4.8
Quantile = 50%
γ = 4.8
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust


Lognormal distribution: γ = 4.8
Quantile = 75%
γ = 4.8
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust


Lognormal distribution: γ = 4.8
Quantile = 90%
γ = 4.8
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
0
5
10
15
20
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust


γ = 8.9
Gamma distribution: γ = 8.9
Quantile = 50%
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust


γ = 8.9
Gamma distribution: γ = 8.9
Quantile = 75%
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust


γ = 8.9
Gamma distribution: γ = 8.9
Quantile = 90%
Sample size
(25,25)
(50,50)
(100,100)
(500,500)
Rejection rates (%)
20
15
10
5
0
Welch U
hsheather
bofinger
chamberlain
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
hsheather
bofinger
chamberlain
epanechnikov
epan2
biweight
cosine
gaussian
parzen
rectangle
triangle
bootstrap 20
bootstrap 100
fitted
residual
kernel
fitted
kernel
iid
robust

RESULTS WITH COVARIATE

- **In general**: similar results
- **Most skewed gamma distribution**: improved results with a strong covariate (however, results were quite awful to begin with)

POWER?

CONCLUSION

- **Bootstrap** is the only method with consistently good performance
- **Gamma distribution creating more problems** than lognormal
- Rejection rates **90% percentile > 75% percentile > 50% percentile**
- **Increasing skewness** -> higher rejection rates